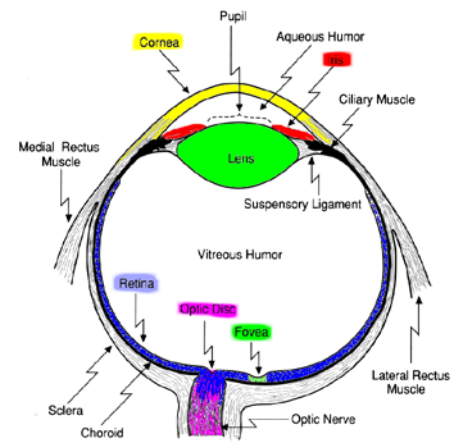


# 16. Bildaufnahme

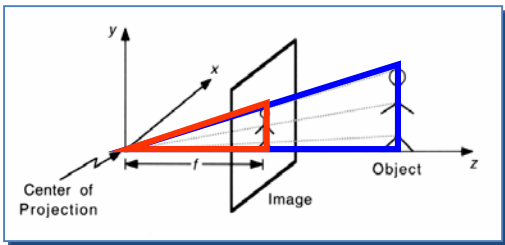
## Das menschliche Auge

Das Auge des Menschen ist ein Organ, welches auf Lichtreize reagiert und für den Menschen das Sehen ermöglicht. Erreicht sichtbares Licht das Auge, kann dieses Informationen über die Umgebung wahrnehmen. Verschiedene physiologische Bausteine sind am Sehen beteiligt und werden unter dem Begriff „visuelles System“ zusammengefasst. Der Vorgang des Sehens funktioniert folgendermaßen: Die Linse des Auges dient zur Fokussierung des Abbildes der Umgebung, welches auf eine lichtempfindliche Membran (Retina/Netzhaut) an der Rückseite des Auges seitenverkehrt und auf dem Kopf stehend projiziert wird. Die Netzhaut besteht aus Fotorezeptoren, welche durch Lichtreize angeregt werden, neuronale Impulse generieren und diese zu verschiedenen Hirnbereichen für die Verarbeitung weiterleiten. Chemische und elektrische Vorgänge sind an der Generierung des neuronalen Impulses beteiligt. Die Retina besteht aus mehreren Schichten von Neuronen, die durch Synapsen miteinander verbunden sind. Bei den lichtempfindlichen Fotorezeptorzellen können zwei Typen unterschieden werden: Stäbchen und Zäpfchen. Stäbchen ermöglichen das Schwarz-Weiß-Sehen, während Zäpfchen für die Wahrnehmung von Farbe verantwortlich sind. Die Fovea centralis liegt im Zentrum des Gelben Fleckes und stellt den Bereich des schärfsten Sehens dar. 50% der Nervenfasern des Sehnervs leiten Impulse der Fovea an das Gehirn weiter, die anderen 50% der Fasern Impulse der restlichen Rezeptoren in der Netzhaut.



## Perspektivische Projektion

Das Modell einer Lochkamera beschreibt die mathematische Beziehung zwischen einem 3D Punkt und dessen Projektion auf eine Bildebene einer idealen Lochkamera. Die Kamerablende wird ohne Verwendung einer Linse zur Fokussierung von Licht verwendet und durch einen Punkt beschrieben. Dies ist das einfachste Bildaufnahmegerät, das die Geometrie der perspektivischen Projektion genau wiedergibt. Das Modell beinhaltet keine geometrischen Verzerrungen oder Ungenauigkeiten von unfokussierten (unscharfen) Objekten, welche durch Linsen und endlich dimensionierte Blenden hervorgerufen werden. Daher wird das Modell nur für eine vereinfachende Annäherung der echten Abbildung einer 3D Szene auf ein 2D Bild herangezogen. Zur Vereinfachung der Herleitung der Gleichung zur perspektivischen Projektion werden folgende Annahmen getroffen (siehe Abbildung links):

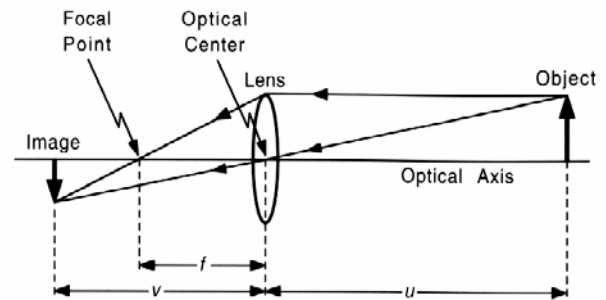


- (1) Das Zentrum der Projektion stimmt mit dem Ursprung des Kamerakoordinatensystems überein.
- (2) Die Kameraachse (optische Achse) ist an der z-Achse der Kamera ausgerichtet.
- (3) Die Bildebene ist im Vordergrund des Zentrums der Projektion, um eine Bildumkehrung zu vermeiden. Das Modell besteht aus einer Ebene (Bildebene) und einem 3D Punkt  $O$  (Zentrum der Projektion). Die fokale Länge  $f$ , die Entfernung zwischen der Bildebene und dem Zentrum der Projektion  $O$  stellt die Brennweite dar (bei digitalen Kameras ist das die Entfernung der Linse zum CCD Chip/Film). Die optische Achse ist jene Gerade die durch  $O$  führt die normal auf die Bildebene steht. Folgende ähnliche Dreiecksbeziehungen bestehen dabei:  $\frac{y}{Y} = \frac{f}{Z}$ .  $y$  stellt die Projektion des Objektes in der Bildebene dar,  $Y$  die Objektgröße und  $Z$  die Entfernung des Objektes zu  $O$ . Folgende ähnliche Zusammenhänge können daraus abgeleitet werden:  $x = \frac{f}{Z} X$ , und  $y = \frac{f}{Z} Y$ , welche die Gleichungen der perspektivischen Projektion darstellen. Diese Projektion ist nicht linear (die Entfernung zu einem Objekt ist umgekehrt proportional zu dessen Bildgröße), ist eine 1:n Abbildung (mehrere 3D Punkte können auf denselben Punkt abgebildet werden), 3D Linien werden auf 2D Linien projiziert (außer sie sind parallel zur optischen Achse), Entfernungen und Winkel bleiben nicht erhalten und parallele Linien werden nicht auf parallele Linien projiziert (sofern diese nicht parallel zur Bildebene liegen). Im Allgemeinen geht die 3D Informationen in einer 2D Projektion für immer verloren.

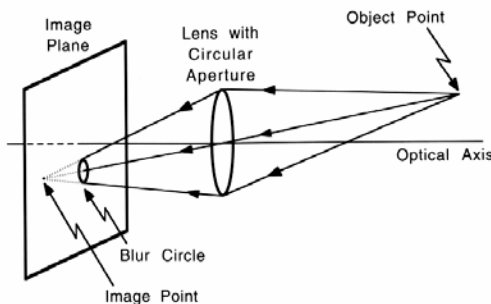
## Linsen

Während die einfache Geometrie der Lochkamera sehr anschaulich ist, hat die Kamera selbst in der Praxis keine Bedeutung. Um eine scharfe Projektion zu erzielen, benötigt man eine möglichst kleines Loch (Lochblende), das aber wenig Licht durchlässt und daher sehr lange Belichtungszeiten benötigt werden. In der Realität verwendet man optische

Linsen und Linsensysteme, deren Abbildungsverhalten besser, aber auch komplizierter ist. Man bedient sich zunächst eines einfachen Modells der „dünnen Linse“, das lediglich die Lochblende durch eine symmetrische und unendlich dünne Linse ersetzt, d.h., jeder Lichtstrahl, der in die Linse fällt, wird an einer virtuellen Ebene in der Linsenmitte gebrochen. Daraus ergibt sich die gleiche Abbildungsgeometrie wie bei einer Lochkamera. Parallele Strahlen, die in Richtung der optischen Achse einfallen, werden im Brennpunkt, der die Entfernung  $f$  zur Linse hat, fokussiert. Liegt ein Objekt in einer endlichen Entfernung  $u$ , werden dessen Lichtstrahlen zu einem Bild hinter dem Brennpunkt in der Entfernung  $v$  fokussiert. Die zur Linsenachse senkrecht stehende Ebene wird Bildebene genannt. Die Beziehung zwischen der Entfernung  $u$  vom Objekt zur Linse und der Entfernung  $v$  von der Linse zur Bildebene wird durch die **einfache Linsengleichung** beschrieben:  $\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$ . Diese besagt, dass Objekte, die sehr weit weg sind ( $u = \infty$ ,  $v=f$ ), im Brennpunkt scharf abgebildet sind, alle Objekte, die näher sind, dahinter.



## Tiefenschärfebereich (DOF)

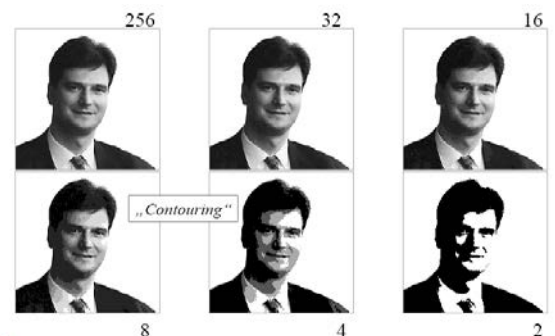


Theoretisch sind auf Grund der einfachen Linsengleichung immer nur Objekte mit einer bestimmten Entfernung zu Kamera scharf abgebildet, alle anderen unscharf. In der Praxis werden jedoch Objekte in einem gewissen Tiefenschärfebereich (engl.: Depth Of Field (DOF)) scharf abgebildet. Der DOF ist somit die Entfernung zwischen dem nächsten und weitesten entfernten Objekt einer Szene, welches scharf in einem Bild dargestellt werden kann. Der Tiefenschärfebereich eines Objekts wird durch die Sensorauflösung und die Linsengröße bzw. Blendenöffnung beeinflusst. Die Blende eines optischen Systems ist eine ringförmige Öffnung vor der Linse, welche den Kegelwinkel des Strahlenbündels

bestimmt, welches auf die Bildebene trifft (je kleiner die Blende, desto weniger Strahlen treffen die Bildebene und desto dunkler ist das Bild für eine gegebene Belichtungszeit, aber desto größer ist der DOF). Der Tiefenschärfebereich ist durch die Größe der Linse bestimmt, je größer die Linse, desto mehr Strahlen eines einzigen Objektpunktes gelangen auf die Bildebene (siehe Abbildung). Ist diese nicht im richtigen Abstand  $v$  zur Linse (zu weit vorne oder hinten) fallen die Strahlen nicht mehr in einem Punkt zusammen, sondern bilden einen Kreis (Blur Circle) auf der Bildebene. Benachbarte Objektpunkte bilden überlagerte Kreise auf der Bildebene – das Bild wird unscharf. Der Bereich der scharf abgebildet wird, ist somit abhängig von der Größe der Unschärfekreise, die wiederum von der Größe der Linse abhängen (d.h. eine größere Linse führt somit zu einem kleineren DOF). Weiters spielt auch die Sensorauflösung eine Rolle, je höher die Auflösung, desto besser können auch die Unschärfekreise abgebildet werden (DOF wird kleiner). Auch die Brennweite hat einen Einfluss auf den DOF: je größer die Brennweite der Linse (Zoom), desto kleiner der DOF.

## Radiometrie

Radiometrie bezeichnet Methoden zur Messung elektromagnetischer Strahlung (inklusive des sichtbaren Lichtes). Die radiometrische Beziehung zwischen der Welt und deren Projektionen wird durch die Menge des gesamten reflektierten Lichtes eines Oberflächenpunktes beschrieben, welche auch als Strahldichte (engl. radiance) bezeichnet wird. Die Menge an Licht, die von diesem Punkt auf das Bild projiziert wird, wird Bestrahlungsstärke (engl. irradiance) oder Helligkeit genannt. Diese radiometrischen Größen charakterisieren die Strahlung aller Frequenzen, betrachtet man jeden Frequenzbereich getrennt, nennt man dies Spektrum. Das Strahlungsaufkommen einer Oberfläche wird spektrale Bestrahlungsstärke (engl. spectral irradiance) genannt.



## Sampling

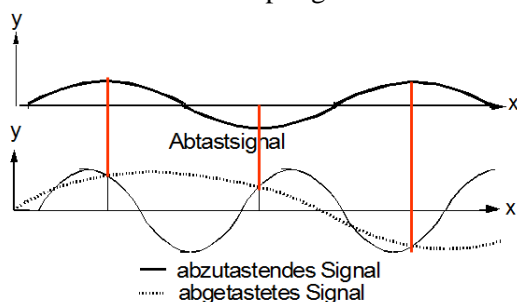
Die Bildauflösung ist ein Sammelbegriff für den Detailreichtum eines Bildes und wird von der *radiometrischen*-, *Sensor*-, *räumlichen*- und *zeitlichen Auflösung* bestimmt. Die **radiometrische** Auflösung bestimmt, wie fein ein System Helligkeitsunterschiede unterscheidet, und wird durch die Anzahl von Grauwertstufen oder Bits beschrieben (typischerweise 8 Bit oder  $2^8 = 256$  Grauwertstufen). Der Mensch kann ca. 120 verschiedene Grauwerte erkennen, somit reichen 256 Graustufen aus, um ein Bild realitätsnah zu erzeugen. Werden weniger Graustufen verwendet, sinkt der Speicherbedarf, aber auch die Qualität und Contouringeffekte entstehen.

Die **Sensorauflösung** steht für die Anzahl der Bildelemente (Picture-elements = Pixel) in digitalen Bildern, d.h. die Anzahl der Pixel pro Zeile (Breite) mal Anzahl der Pixel pro Spalte (Höhe), z.B. 640 x 480, und wird üblicherweise in Megapixel angegeben. Eine andere Möglichkeit, die Sensorauflösung zu beschreiben, ist Pixel per Längeneinheit (engl. dots per inch (dpi) oder pixels per inch (ppi)). Diese ist aber nur für Flachbettscanner gültig, da dort der Abstand zwischen Objekt und Sensor bekannt ist, für Kameras gilt dies nicht.

Diese sogenannte **räumliche** (geometrische) Auflösung kennzeichnet allgemein die Fähigkeit eines Sensorsystems, Signale von benachbarten Objektstrukturen getrennt zu erfassen. Es werden verschiedene Maße für die Auflösung benutzt, in der Optik und Photographie werden Linien je Millimeter (L/mm) verwendet, die angeben, wie nahe beieinander liegende Linien in einem Bild aufgelöst werden können. Sie wird also durch die minimale Entfernung zwischen zwei benachbarten Merkmalen oder die minimale Größe eines Merkmals definiert.

Die **zeitliche** Auflösung gibt an, wie viele Bilder pro Sekunde aufgenommen werden (z.B. 25 frames per second).

In digitalen Kameras wird anhand der zur Verfügung stehenden Auflösungen eine analog/digital Wandlung im Raum, in der Zeit und im Spektrum durchgeführt. Die Auflösungsgrenze wird durch das *Nyquist-Shannon Sampling Theorem* berechnet. Unter Sampling versteht man den Umwandlungsprozess von einem Signal (z.B. eine kontinuierliche



Funktion in Zeit oder Raum) zu einer numerischen Folge (eine diskrete Funktion in Zeit oder Raum). Das Theorem besagt, dass ein bandbegrenzt analoges Signal perfekt von einer unendlichen Folge von Samples rekonstruiert wird, wenn die Samplingrate  $2 \cdot B$  Samples pro Sekunde (engl. samples per second) übersteigt.  $B$  steht für die höchstvorherrschende Frequenz im Originalsignal. Das heißt um ein Signal  $f(t)$  exakt wiederzugewinnen, muss man  $f(t)$  mit einer Rate abtasten, die **größer als das Doppelte** seiner höchsten Frequenzkomponente ist (f abtast  $> 2 \cdot f_{\text{maxSignal}}$ ).

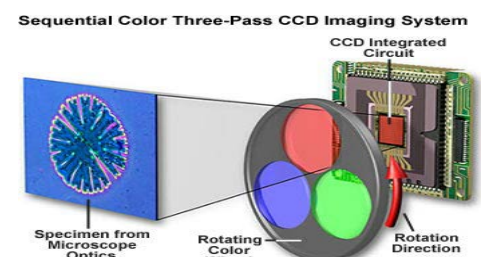
## Bildsensoren

Ein Bildsensor ist ein Bauelement, das ein optisches Bild in ein elektronisches Signal umwandelt. Es sind zwei Arten von Pixelsensoren (*digitale Charge-Coupled Device (CCD)* oder *Complementary Metal-Oxide-Semiconductor (CMOS)*) in Verwendung. CCD Sensoren bestehen aus einer Matrix (engl. array) lichtempfindlicher Fotodioden, kleinsten MOS-Kondensatoren, Feldeffekttransistoren, Steuerleitungen und Leitungspfaden. Bei einer gegebenen Belichtungszeit (engl. exposure time) baut sich, während Licht auf den CCD Sensor trifft, elektrische Ladung auf, die von einem Kondensator, der zur Fotodiode parallel geschaltet ist, gespeichert wird. Anschließend werden Zeile für Zeile die Spannungen der Kondensatoren ausgelesen, die anschließend entweder innerhalb der Kamera oder außerhalb in digitale Information umgewandelt werden. CCD Kameras arbeiten wie auch analoge Kameras nach dem Prinzip des Flächenintegrals, und sind daher linear. Ein CMOS Bildaufnahmechip hingegen besitzt aktive Pixelsensoren, jedes Pixel ist individuell auslesbar, die Spannung, die der Helligkeit entspricht, liegt immer an. Dieser Sensor hat durch die unabhängige Betrachtung der Nachbarn jedes Pixels einen nicht-linearen (logarithmischen) Output, der im Wesentlichen dem Prinzip der menschlichen Bildaufnahme im Auge entspricht. CMOS haben im Vergleich zu CCD Sensoren einen geringeren Energieverbrauch, schnellere Lesegeschwindigkeit und sind in der Herstellung billiger. In der Masse der industriellen Applikationen werden Kameras mit CCD-Sensoren verwendet. Sie sind rauscharm, lichtempfindlich, haben eine hohe Uniformität und sind aufgrund Ihrer linearen Charakteristik geeignet für präzise Messapplikationen.

## Farbe auf CCD Sensoren

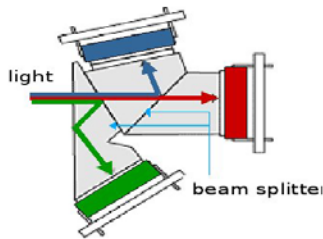
Bildsensoren können nur Helligkeit messen. Um Farbinformationen zu erlangen, muss man die auf die lichtempfindliche Sensorfläche auftreffenden Photonen einen Farbfilter passieren lassen. Digitale Farbkameras können Farbe auf drei verschiedenen Wegen erfassen: Field Sequential Technik, 3 Chip und Color Filter Array (CFA).

Die **Field Sequential Technik** wurde in den frühen 1900er Jahren durch Prokudin Gorskii erfunden, der die Farbinformation der Szene dadurch gewann, dass er mithilfe von Rot-, Grün- und Blau-Filtern drei Schwarz-Weiß Bilder nacheinander aufnahm und auf Glasplatten entwickelte. Durch den erneuten Einsatz von Farbfiltern war er in der Lage, diese Bilder mithilfe eines Lichtprojektionssystems (Lantern Projektor) in Farbe darzustellen. Jede Bewegung in der Bildaufnahme resultiert aber in



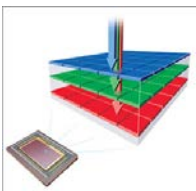
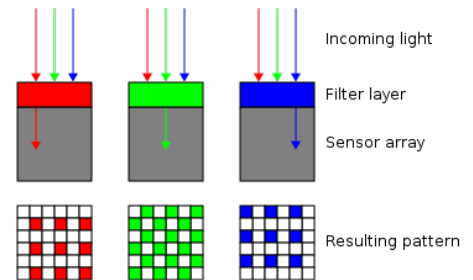


„Geisterbildern“, da die roten, grünen und blauen Bilder der Objekte zu verschiedenen, hintereinander folgenden Zeitpunkten gemacht werden. Diese Art der Farbaufnahme wird heute noch vor allem bei Mikroskopen verwendet, da sich dort die Szene üblicherweise nicht bewegt.



Durch die Verwendung eines strahlenbrechenden Prismas kann eine korrekte Farbtrennung erzielt werden, da das Eingangssignal in einen roten, grünen und blauen Anteil aufgespalten wird. In **3 Chip-CCD Kameras** ist jeder dieser drei CCDs so angeordnet, dass er genau einen Farbkanal aufnimmt. Das Problem dabei ist es, die CCD Chips so anzuordnen, dass ein eintretender Strahl genau auf denselben Pixel in allen 3 CCD Chips fällt, denn nur so kann die Farbtreue gewährleistet werden. Diese Technik wird ausschließlich für Videokameras und Camcorder verwendet.

Die meistverwendete Technik zur Farbaufnahme mit CCDs sind **Color Filter Array (CFA)** oder Color Filter Mosaic (CFM), bei der ein Mosaik aus winzigen Farbfilttern, die über jedes Pixel eines Bildsensors platziert werden, um Farbinformation aufzunehmen, verwendet wird. Die am häufigsten verwendeten CFAs, Bayer Filter genannt, bestehen aus 50% grünen, 25% roten und 25% blauen Filtern und sind daher an die Physiologie des menschlichen Auges angepasst. Der Mensch ist für grünes Licht am empfindlichsten und der Wellenbereich für Grün liegt zwischen Blau und Rot. Der rohe (engl. raw) Output einer Bayer-Filter Kamera wird als Bayer Muster Bild (engl. Bayer pattern image) bezeichnet. Jedes Pixel nimmt nur eine der drei Farben auf, somit muss die tatsächlich Farbe des Pixels, mit Hilfe von Interpolationsalgorithmen berechnet werden, um die fehlenden Rot, Grün und Blauanteile je Farbkanal für jeden fehlenden Punkt aus dessen Nachbarn zu berechnen (engl. demosaicing). Bei der Farbinterpolation geht man davon aus, dass es im Bild zwischen zwei benachbarten Pixeln gleicher Farbe nur zu geringen Farbunterschieden kommt und somit die Grauwerte der Pixel stochastisch nicht unabhängig voneinander sind. Daher besitzt der Bayer Sensor unter der Betrachtung eines artefakt-freien Bildes streng genommen nur ein Drittel der angegebenen geometrischen Auflösung. Der Rechenvorgang kann innerhalb der Kamera gemacht werden, z.B. durch Konvertierung des Raws in ein JPEG oder TIFF Format, oder außerhalb der Kamera, durch die direkte Verwendung der Rohdaten des Sensor in einem Bildanzeigeprogramm (wie zum Beispiel nomacs, [www.nomacs.org](http://www.nomacs.org)).



Der Foveon X3 Sensor ist eine Kombination der 3CCD und der CFA Technik, verwendet ein Array von Fotoelementen, welches jeweils aus drei vertikal gestapelten Fotodioden besteht und in einem zweidimensionalen Raster angeordnet ist. Jeder dieser gestapelten Fotodioden reagiert auf verschiedene Wellenlängen des Lichtes, d.h. jede hat eine unterschiedliche spektrale Sensitivitätskurve. Das Signal der drei Fotodioden wird anschließend verarbeitet und resultiert in Daten, welche drei additive Primärfarben rot, grün und blau beinhalten und das Farbbild fehlerfrei darstellen kann.

## Ideale vs. Reale Kamera

Reale Kameras weichen von idealen Kameramodellen ab, da sie Linsen mit optischen Phänomenen und nicht-ideale Bildsensoren mit Chipartefakten verwenden. Einige dieser Störungen sind:

Die *optische Aberration* tritt auf, wenn Licht von einem Punkt auf einem Objekt nicht in einem einzigen Punkt nach der Transmission durch die Linse zusammenfällt.

*Chromatische Aberration* ist eine Art von Störung, bei der die Linse nicht im Stande ist, alle Farben im gleichen Konvergenzpunkt zu fokussieren. Diese Aberration tritt auf, da Linsen unterschiedliche Brechungsindizes für verschiedene Wellenlängen des Lichtes aufweisen. Chromatische Aberration manifestiert sich durch „Farbränder/-säume“ entlang der Abgrenzung zwischen dunklen und hellen Bildteilen.

Die *sphärische Aberration* tritt auf, da Kugeloberflächen nicht die ideale Form für Linsen sind, aber beim Schleifen von Glas die einfachste Form darstellen. Diese Aberration führt dazu, dass Strahlen, die weiter am Rand der Linse auftreffen in einem etwas anderen Punkt fokussiert werden, als Strahlen die nahe der Achse liegen.

*Linsen Verzerrung* ist eine Abweichung einer geradlinigen Projektion, gerade Linien in einer Szene bleiben nicht gerade im Bild, diese können nach außen (tonnenförmig) oder nach innen (kissenförmig) verzerrt sein.

